The EMBO Journal (2006) 25, 1295–1304 | © 2006 European Molecular Biology Organization | All Rights Reserved 0261-4189/06
www.embojournal.org

THE EMBO JOURNAL

Retroviral DNA integration: reaction pathway and critical intermediates

Min Li[1], Michiyo Mizuuchi[1], Terrence R Burke Jr[2] and Robert Craigie[1,*]

[1]Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA and [2]Laboratory of Medicinal Chemistry, Center for Cancer Research, NCI-Frederick, National Cancer Institute, National Institute of Health, Frederick, MD, USA

The key DNA cutting and joining steps of retroviral DNA integration are carried out by the viral integrase protein. Structures of the individual domains of integrase have been determined, but their organization in the active complex with viral DNA is unknown. We show that HIV-1 integrase forms stable synaptic complexes in which a tetramer of integrase is stably associated with a pair of viral DNA ends. The viral DNA is processed within these complexes, which go on to capture the target DNA and integrate the viral DNA ends. The joining of the two viral DNA ends to target DNA occurs sequentially, with a stable intermediate complex in which only one DNA end is joined. The integration product also remains stably associated with integrase and likely requires disassembly before completion of the integration process by cellular enzymes. The results define the series of stable nucleoprotein complexes that mediate retroviral DNA integration.

The EMBO Journal (2006) **25,** 1295–1304. doi:10.1038/sj.emboj.7601005; Published online 16 February 2006
Subject Categories: genome stability & dynamics
Keywords: integrase; integration; retrovirus; transposition

Introduction

Integration of retroviral DNA into the genome of the host cell is an essential step in the viral replication cycle (Brown, 1997; Craigie, 2002). The viral DNA made by reverse transcription is initially blunt-ended and is processed by the viral integrase protein to remove two nucleotides from each 3′-end. This 3′ end-processing step exposes the conserved CA dinucleotides that ultimately define the ends of the integrated viral DNA. Integration occurs by a pair of transesterification reactions, catalyzed by integrase, in which the 3′ hydroxyl groups of the processed viral DNA ends attack a pair of phosphodiester bonds in the target DNA. In the case of HIV-1, the sites of attack on the two target DNA strands are separated by 5 bp. The 3′ ends of the viral DNA are joined to target DNA, whereas the 5′ ends of the viral DNA are unjoined in the resulting integration intermediate. Repair of the intermediate by cellular enzymes completes the integration process.

Viral DNA made by reverse transcription is part of a large nucleoprotein complex, termed the preintegration complex (PIC), that is derived from the core of the infecting virion (Bowerman *et al*, 1989). Integrase and other viral and cellular proteins are stably associated with the viral DNA in the PIC. It is not known if components of the PIC other than integrase remain stably associated until the PIC is transported to the nucleus and integrates the viral DNA, or whether they are jettisoned along the way. PICs isolated from infected cells efficiently integrate their DNA *in vitro* with the full fidelity and all the hallmarks of integration *in vitro* (Brown *et al*, 1987; Ellison *et al*, 1990; Farnet and Haseltine, 1990). However, PICs are present in low abundance in cell extracts and this limits the biochemical studies that can be attempted.

Both the 3′ processing and DNA strand transfer activities of HIV-1 integrase can be recapitulated *in vitro* with DNA substrates that mimic the viral DNA ends (Bushman *et al*, 1990; Sherman and Fyfe, 1990; Engelman *et al*, 1991). However, under most reaction conditions the strand transfer products result from a half-site reaction in which only one viral DNA end is joined to one strand of target DNA, rather than concerted integration of a pair of viral DNA ends as occurs *in vivo*. The reasons for this poor *in vitro* fidelity are unclear, but modifications of reaction conditions have boosted the efficiency of concerted integration to a level at which it can be studied biochemically (Hindmarsh *et al*, 1999; Sinha *et al*, 2002; Li and Craigie, 2005; Sinha and Grandgenett, 2005). Overall, the half-site integration is more tolerant of a broader range of reaction conditions than the concerted reaction. In addition, certain DNA substrate properties favor one outcome over the other. For example, preprocessed DNA lacking the two nucleotides normally removed by integrase is a more efficient substrate for the half-site reaction than is blunt-ended viral DNA, whereas blunt-ended DNA favors the concerted over the half-site reaction (Li and Craigie, 2005). Also, DNA substrates shorter than 20 bp efficiently perform half-site integration while several hundred base pairs are required for maximally efficient concerted integration (Li and Craigie, 2005). The reasons for the different requirements for concerted versus half-site integration are not well understood.

The chemical steps of retroviral DNA integration closely parallel those of 'classical' transposons such as bateriophage Mu and insertion elements (Mizuuchi, 1992, 1997). For some of these transposons, the organization of the nucleoprotein complexes that carry out transposition have been extensively studied. Mu transposase forms a stable tetramer with the Mu DNA ends and only two of the active sites within the tetramer are involved in catalysis (Chaconas and Harshey, 2002). In contrast, Tn5 transposase functions as a dimer (Davies *et al*, 2000). In the case of retroviral DNA integration,

*Corresponding author. Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, Bldg 5, Room 301, National Institutes of Health, 5 Center Drive MSC 0560, Bethesda, MD 20892, USA. Tel.: +301 496 4081; Fax: +301 496 0201; E-mail: bobc@helix.nih.gov

Received: 7 October 2005; accepted: 25 January 2006; published online: 16 February 2006

the organization and stoichiometry of nucleoprotein complexes along the reaction pathway is largely unknown.

Structures of all three domains of HIV-1 integrase have been determined (Dyda *et al*, 1994; Lodi *et al*, 1995; Cai *et al*, 1997), as have several two-domain structures (Chen *et al*, 2000a; Wang *et al*, 2001). Domain structures of the closely related SIV and RSV integrase have also been solved (Bujacz *et al*, 1995; Chen *et al*, 2000b; Yang *et al*, 2000). However, no structures have been solved in the presence of DNA substrates. Integrase binds DNA non-specifically in simple DNA-binding assays and the protein and DNA aggregate under reaction conditions. Consequently, physical information on the nucleoprotein complexes that mediate retroviral DNA integration is limited. Kinetic analysis combined with atomic force microscopy support the idea that the disintegration reaction of avian sarcoma virus integrase is mediated by a tetramer (Bao *et al*, 2003). However, because this substrate contains only a single viral DNA end, it is not clear whether this can be extrapolated to concerted integration of pairs of viral DNA ends. Crosslinked tetramers of HIV-1 integrase have been reported to carry out concerted integration (Faure *et al*, 2005). However, the integrase preparation used for these studies was inactive with the normal blunt-ended DNA substrate for concerted integration. Improvements in the reaction conditions for concerted integration *in vitro* (Li and Craigie, 2005) have allowed us to determine the stable complexes of integrase and DNA that participate in concerted DNA integration. We show that under conditions that promote concerted DNA integration, HIV integrase forms highly stable complexes with viral DNA that are apparently as stable as the association of integrase with viral DNA in the PIC. Each step of the integration reaction occurs within these stable complexes. The results define the active unit of integrase and establish that retroviral DNA integration occurs in the context of a series of highly stable nucleoprotein complexes. These complexes progress from initial synapsis of a pair of viral DNA ends to the integration intermediate in which viral and host DNA are covalently joined.

Results

The products of concerted DNA integration are stably associated with integrase

The reaction system for concerted DNA integration was essentially as described by Li and Craigie (2005), except that substrate with only a single viral DNA end was used to simplify the analysis of integration products; concerted integration linearizes the circular target DNA by integration of a pair of viral DNA ends, whereas the half-site reaction tags the circular target DNA with a single viral DNA end (Figure 1A). The deproteinized products of such a reaction, after separation by gel electrophoresis, are shown in Figure 1B, lane 1. When the same reaction products are subjected to electrophoresis without deproteinization (lane 2), the band corresponding to the concerted integration products disappears, and a prominent new band with slower electrophoretic mobility is observed. As we show below, this strand transfer complex (STC) contains viral DNA, target DNA and integrase. In early experiments, the STC ran as a smear from the top of the gel that ended abruptly at the position of the STC shown in Figure 1B. This smearing, which is likely the result of interaction between the STC and the gel matrix, or less stable interactions among STCs mediated by other integrase protomers, was minimized by modifications to the electrophoresis conditions as noted in Materials and methods. The STC is stable to high ionic strength, challenge with heparin and to detergent (lanes 3–5). In contrast to the products of concerted integration, the half-site integration products are not stably associated with integrase and migrate at the same position with or without deproteinization.

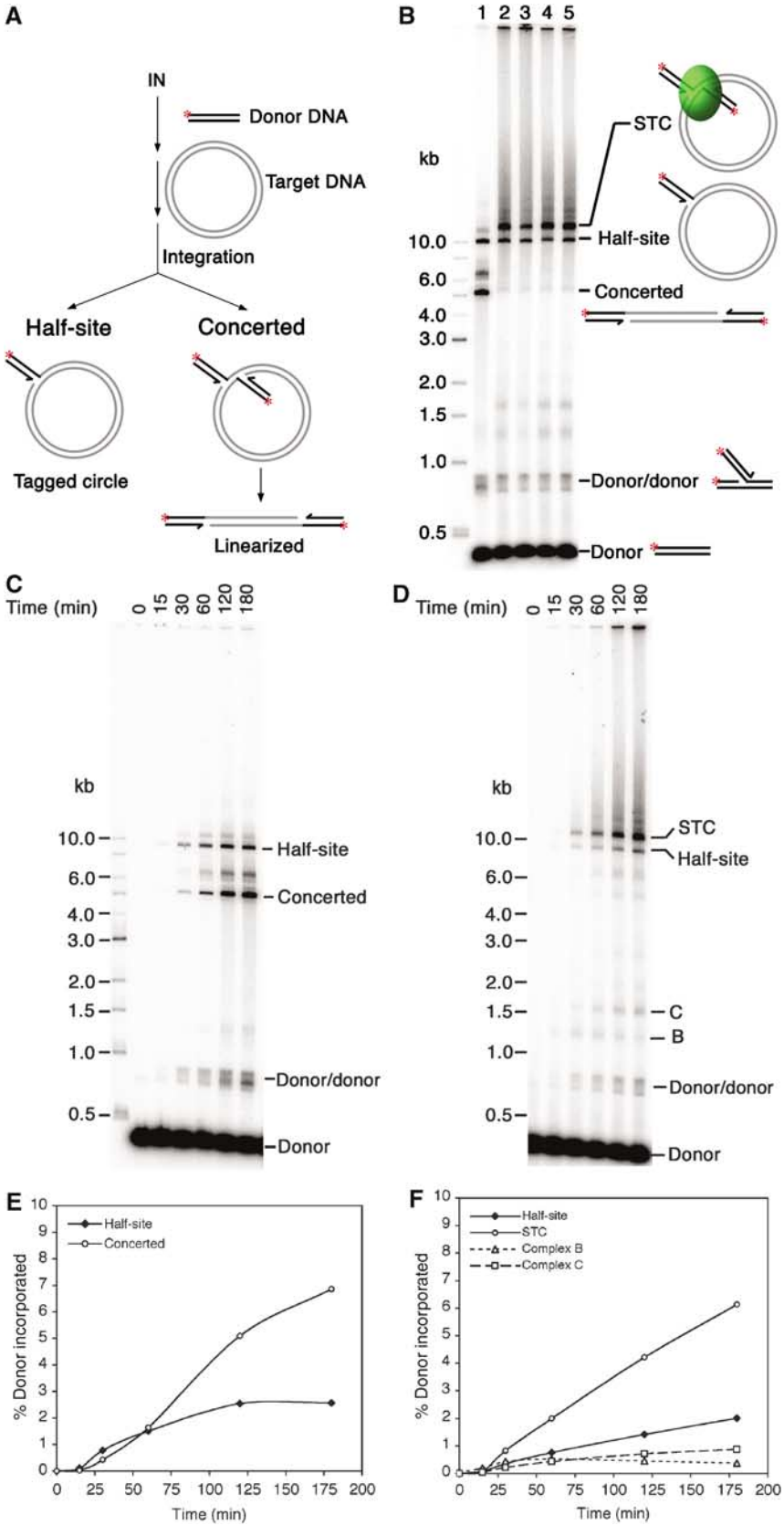
As expected, comparison of the same samples subjected to electrophoresis with and without deproteinization revealed that the kinetics of formation of concerted integration products paralleled formation of the STC (Figure 1C and D). Quantitation of the data (Figure 1E and F) confirms that, as expected, the percentage of donor DNA incorporated into the STC closely correlates with the percentage of donor DNA incorporated into the concerted integration product upon deproteinization. In addition to the STC, two other complexes with different electrophoretic mobility are observed (Figure 1D, complexes B and C). Complex B appears early in the time course and decreases in abundance at later time points. In contrast, the amount of complex C steadily increases with kinetics similar to that of the STC. The identity of these complexes was established as described below.

The two DNA strand transfer steps occur sequentially

We employed two-dimensional gel analysis to determine the DNA species present in each complex. Reaction products were first subjected to electrophoresis in an agarose gel without deproteinization. Integrase was then dissociated from the DNA by soaking the gel in SDS, and the reaction products were then subjected to electrophoresis in a second dimension perpendicular to the first (Figure 2A). The STC, which migrates as a single band in the first dimension, split into two bands in the second dimension. One band has the size expected for the concerted integration product, whereas the other migrates identically to the half-site tagged-circle product in the second dimension. We conclude that some of the STC has both viral DNA ends covalently joined to target DNA, whereas the remainder has only one end joined and the other end is retained in the complex by non-covalent protein–DNA interactions. Consistent with this interpretation, deproteinization of the STC releases some DNA that migrates as a

Figure 1 Concerted integration products are stably associated with integrase. (**A**) Schematic showing the expected integration products after deproteinization. (**B**) The products of an integration reaction were subjected to electrophoresis in an agarose gel after deproteinization (lane 1), after the addition of EDTA (lane 2), EDTA and 600 mM NaCl (lane 3), EDTA and 10 μg/ml heparin (lane 4) and EDTA and 0.1% NP-40 (lane 5). The migration position of linear size markers is shown on the left. (**C, D**) Kinetics of STC formation parallels formation of concerted integration product. Reaction mixtures were incubated for the indicated time and subjected to electrophoresis with (C) or without (D) deproteinization. The band just above the 6 kb marker in panel B, lane 1 and in panel C results from a half-site insertion of another donor DNA molecule into the concerted integration product. (**E, F**) show quantitation of the data in panels C and D, respectively. This quantitation slightly under-represents the abundance of the SSC in panel F because some of the complex runs as a smear. Similarly, in the absence of deproteinization, some of the half-site product in panel F runs as a smear and is slightly under-represented.

free viral DNA end in the second dimension of electrophoresis. To test whether the STC with only one end joined is an intermediate on the pathway to concerted integration, we carried out a time course of two-dimensional gel analysis. Figure 2B shows that the two complexes are similar in abundance at a 0.5 h time point. At 1 h, the complex with two ends joined is more abundant and by 2 h essentially all the STC is converted into this form. We conclude that the



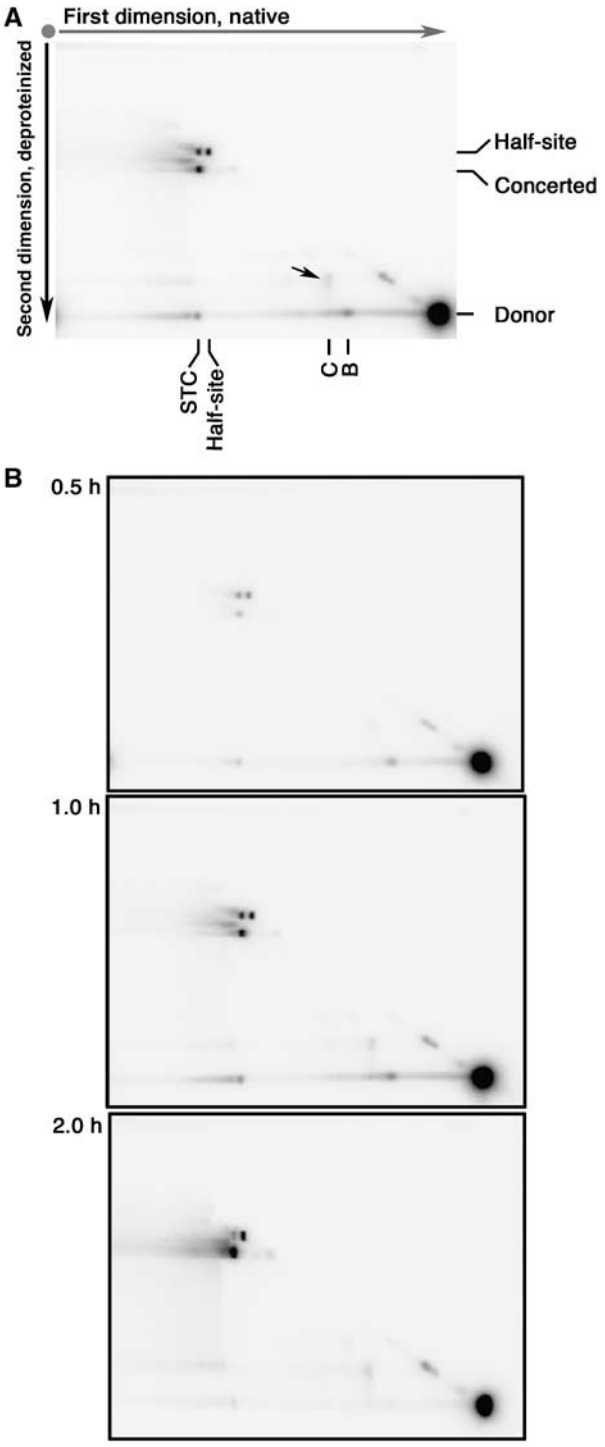


Figure 2 Analysis of the stable complexes by two-dimensional gel electrophoresis. (**A**) The products of a 1 h reaction were subjected to electrophoresis in a native dimension (left to right). The gel was then soaked in SDS to dissociate protein–DNA complexes and then subjected to electrophoresis in a second dimension (top to bottom). (**B**) Time course of two-dimensional gel analysis. Each panel is the same as in (A), except that reactions were incubated for 0.5, 1 or 2 h. Note that the complex with only one viral DNA end joined is chased to concerted product by the 2 h time point. The smear resulting from integration of pairs of viral ends into another viral DNA end is indicated by the arrow. Quantitation of the data shows that the percentage of STC with both viral DNA ends covalently joined at 0.5, 1.0 and 2.0 h is 39, 60 and 89%, respectively.

joining of the two viral DNA ends occurs sequentially and with slow kinetics. It is noteworthy that at the later time point essentially all the STC has two ends joined. Thus, once the correct complex for concerted integration has been assembled, the reaction is highly efficient. Two-dimensional gel analysis also allowed us to determine the composition of complexes B and C first shown in Figure 1. Complex B migrates identically to viral DNA substrate in the second dimension and is therefore a non-covalent complex of viral DNA with integrase. The kinetics of its formation and disappearance are consistent with it being an intermediate on the concerted integration pathway (Figure 1). Complex C runs as a smear in the second dimension (indicated by the arrow in Figure 2A). The similar stability to the STC, the smear in the second dimension of electrophoresis, and the kinetics of its appearance (Figure 1) suggest that complex C is an STC resulting from integration of a pair of viral DNA ends into another linear viral DNA substrate molecule, rather than into the circular target DNA.

Complex B is a stable synaptic complex (SSC) of a pair of viral DNA ends with integrase

Transient accumulation of complex B made it difficult to obtain sufficient complex for analysis. If complex B is being converted to the STC, then a selective inhibitor of strand transfer should cause it to accumulate (Pommier *et al*, 2005). To test this, we included the azido-containing aryl β-diketo acid inhibitor number 11 (Zhang *et al*, 2003), which has high selectivity for inhibition of DNA strand transfer, in the reaction. Figure 3 shows that this inhibitor decreases the formation of STC with concomitant accumulation of complex B. Formation of this SSC is more efficient with longer viral DNA substrate, and with such substrates, high levels of SSC accumulate in the presence of inhibitor (Figure 3). Stable complexes are not formed with DNA that lacks the HIV-1 terminal sequence (Figure 3C).

How many viral end DNA molecules are present in complex B? The strand transfer inhibitor enabled us to accumulate sufficient complex to answer this question by assembling complexes with a mixture of long and short DNA viral DNA ends. If two DNA molecules are present in the complex, three different complexes should be formed that can be distinguished by their different electrophoretic mobility: long plus long, long plus short and short plus short. Figure 4 shows that complex B does indeed split into three bands when assembled with DNAs of two different lengths. As expected, the relative abundance of each species depends on the ratio of long to short DNA. Complex B is therefore an SSC of a pair of viral DNA ends with integrase.

The SSC can contain both processed and unprocessed viral DNA ends

Do both viral DNA ends need to be processed to form the SSC? SSCs were assembled in the presence of inhibitor with a viral DNA substrate with an *Eco*RI restriction site close to the ends of the DNA. The SSC band was cut out of an agarose gel and the DNA was digested with *Eco*RI and subjected to electrophoresis in a denaturing polyacrylamide gel. Bands corresponding to both unprocessed (35 bases) and processed (33 bases) DNA were observed (Figure 5); approximately 30% of the viral DNA ends in the SSC were processed. Therefore, formation of the SSC does not require that both

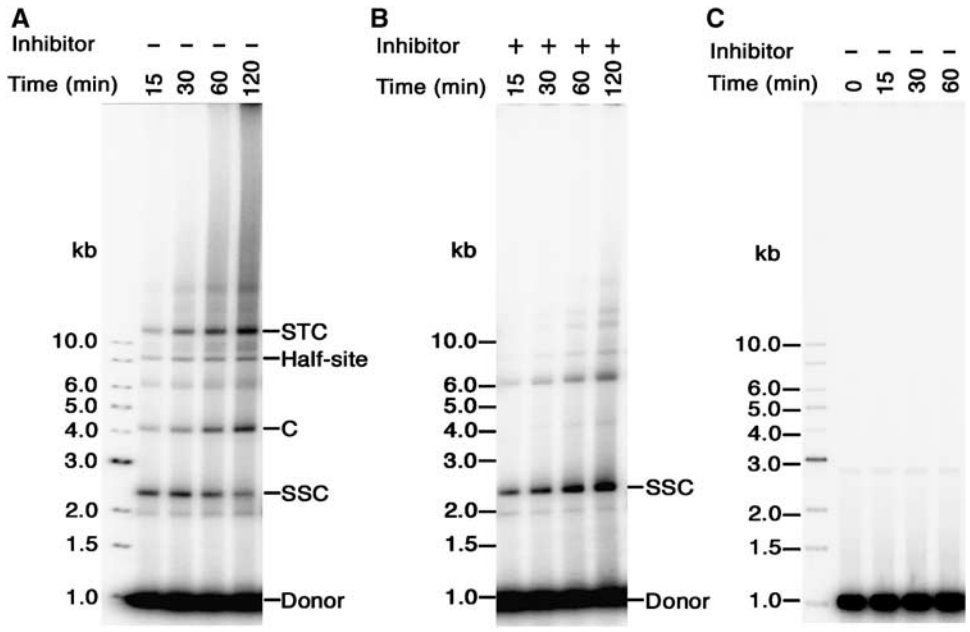


Figure 3 The SSC accumulates in the presence of an inhibitor of DNA strand transfer. Reactions, which were incubated for the indicated times, included a 984 bp viral DNA substrate, which is more efficient for SSC formation than the 355 bp substrate. Products were subjected to electrophoresis in an agarose gel without deproteinization. (**A**), No inhibitor and (**B**) reactions were incubated in the presence of 0.5 mM inhibitor. The band above the 6 kb marker in panel B varies in intensity between experiments and we suspect it represents two SSCs associated with one another. (**C**) Complexes are not formed with a 970 bp substrate that lacks the HIV-1 LTR sequence. The substrate was incubated with integrase and subjected to electrophoresis as for panel A.

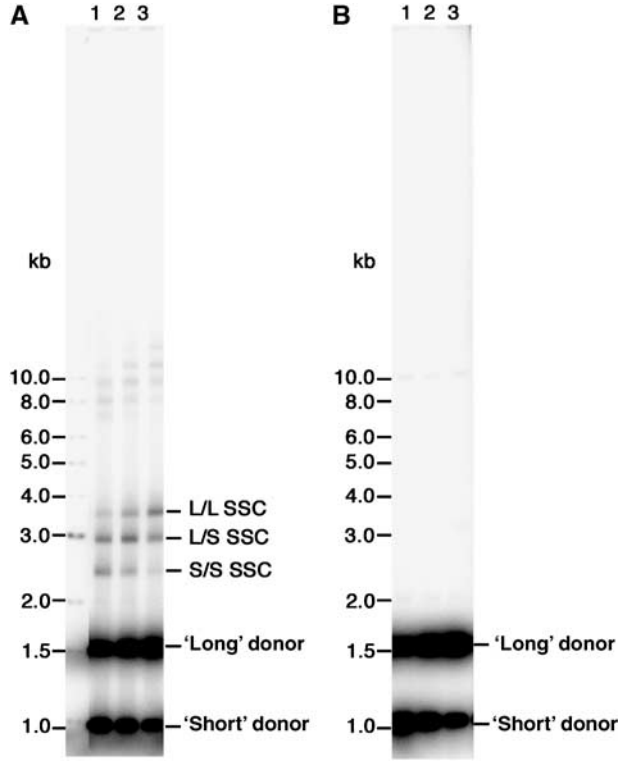


Figure 4 Complex B contains a pair of viral DNA ends. Complexes were assembled with a mixture of 984 and 1513 bp viral DNA ends for 30 min in the presence of inhibitor and subjected to electrophoresis without (**A**) or with (**B**) deproteinization. The ratio of short:-long DNA was, lane 1, 2:1; lane 2, 1:1 and lane 3, 1:2.

ends of the viral DNA be processed and stable association with the target DNA can occur without prior processing of both viral DNA ends.

In the absence of inhibitor, the SSC is a transient intermediate and does not accumulate to a level that permits gel isolation and direct measurement of the extent of processing. However, we argue that once the SSC is formed both ends are eventually processed and processing is not a limiting step in promoting concerted integration. At the earliest time point analyzed (Figure 2B; 0.5 h), approximately 40% of the viral DNA ends in the SSC were covalently joined to target DNA. As sequencing of integration products reveals that integration does not occur in the absence of processing under our reaction conditions (Li and Craigie, 2005), approximately 40% of the viral DNA ends must have been processed at this time point. At the last time point analyzed (Figure 2B; 2.0 h), essentially all the viral DNA ends in the STC had been joined and therefore essentially all the viral DNA must have been processed.

Integrase protects less than 20 bp of terminal viral DNA sequence

In order to determine the extent of terminal viral DNA sequence associated with integrase, SSCs accumulated in the presence of inhibitor were treated by limited digestion with DNase I before separation in an agarose gel. Complexes were excised and the extracted DNA was subjected to electrophoresis in a denaturing polyacrylamide gel electrophoresis (PAGE) gel. Figure 6A shows that protection from DNase I digestion is limited to only 16 nucleotides from the ends of the viral DNA in both the SSC and STC. The extent of protection is similar to that reported for footprinting of RSV integrase on U3 DNA under conditions that promote concerted integration (Vora and Grandgenett, 2001).

Approximately 30% of the viral DNA ends in the SSCs used for the footprinting shown in Figure 6A were processed. Therefore, the footprint must at least reflect the footprint on

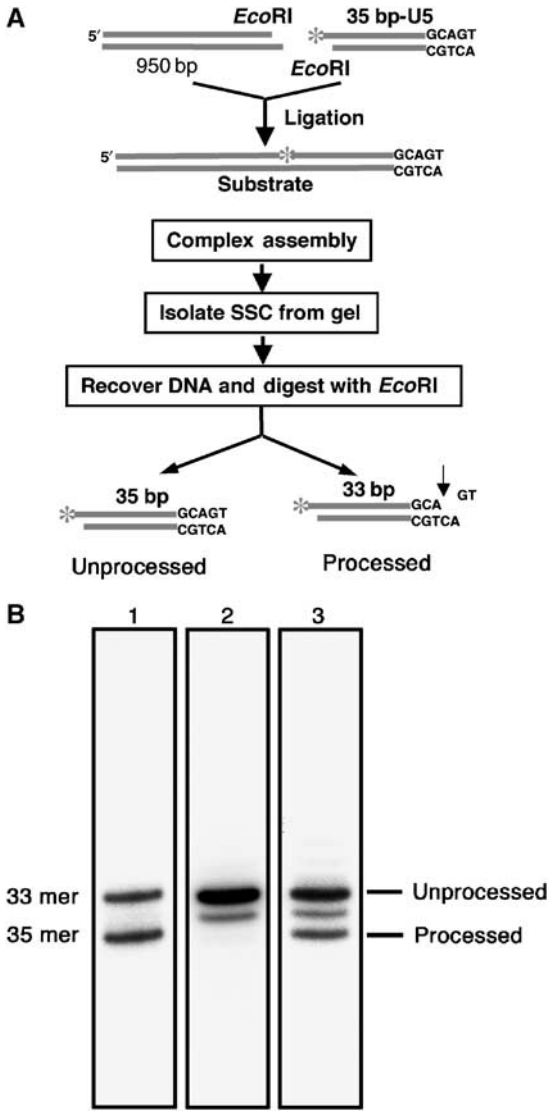


Figure 5 The SSC contains both processed and unprocessed viral DNA. (**A**) In order to enable the base resolution required to monitor 3′ processing, the viral DNA substrate was internally labeled with ^{32}P such that cleavage of unprocessed DNA by *Eco*RI gives a 35-base-labeled fragment and cleavage of processed DNA gives a labeled 33 base fragment upon denaturing gel electrophoresis. (**B**) Lane 1, oligonucleotide markers for the 35 and 33 base cleavage products; lane 2, unreacted viral DNA excised from the gel and processed identically to DNA from the excised SSC and cut with *Eco*RI and lane 3, DNA extracted from SSCs and cut with *Eco*RI. The band migrating at the 34 base position results from a contaminating nuclease activity.

unprocessed viral DNA ends within the SSC. Because this footprint is indistinguishable from that on the viral DNA ends in the STC, we argue that the same protection is exhibited on unprocessed and processed viral DNA ends in the SSC. In order to directly demonstrate that the same footprint is present on processed viral DNA ends in the SSC, we tested whether the SSC could be assembled with processed viral DNA that terminates with 3′ ddA to block DNA strand transfer. Indeed, such processed DNA ends are efficient substrates for assembly of the SSC (data not shown). Footprinting of SSCs assembled with these processed ends confirmed that they exhibit the same footprint as in the SSC containing a mixed population of unprocessed and processed ends (Figure 6B).

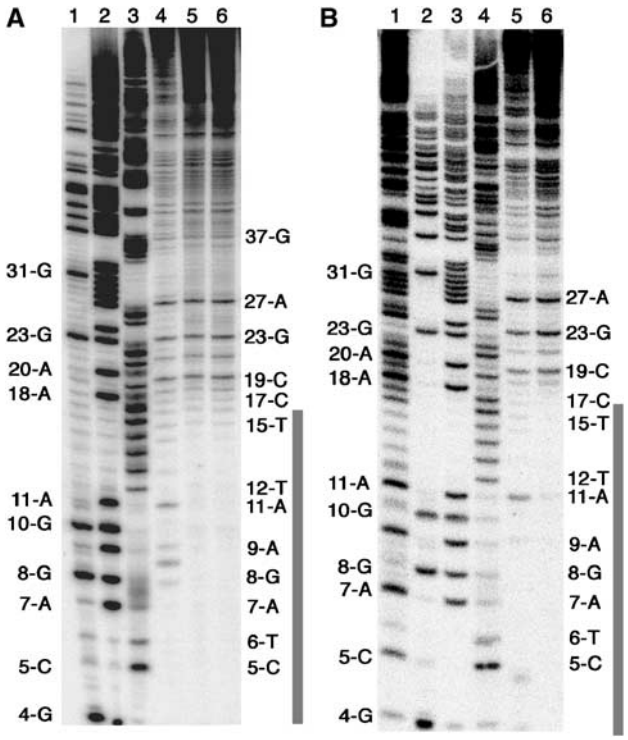


Figure 6 (**A**) Footprint of integrase in the SSC and STC. Complexes were assembled in the presence of inhibitor and treated with limited digestion with DNase I before separation on an agarose gel. After excision of the bands, DNA was extracted and subjected to electrophoresis in a denaturing polyacrylamide gel. Lanes 1, 2 and 3 are G, G + A and C + T markers, respectively, made by Maxam–Gilbert chemical cleavage. The samples shown in lanes 4–6 were processed identically. Lane 4, free DNA; lane 5, STC; and lane 6, SSC. (**B**) Footprint of integrase in SSC complex assembled with processed viral DNA ends with a 3′ ddA that blocks strand transfer. Lanes 1, 2, 3 and 4 are A + C, G, G + A and C + T chemical cleavage markers, respectively. Lanes 5 and 6 are free DNA and SSC, respectively.

A tetramer of integrase is stably associated with the viral DNA ends in the SSC

The multimeric form of integrase in the SSC was determined by crosslinking with the bifunctional crosslinker disuccinimidyl suberate (DSS). SSCs were assembled and crosslinked before electrophoresis in an agarose gel. The SSC band was excised and protein was extracted and subjected to electrophoresis in a denaturing polyacrylamide gel (Figure 7). Control samples of crosslinked integrase in solution displayed a continuous ladder of crosslinked products. In contrast, integrase extracted from the SSC exhibited only two bands corresponding to tetramer and dimer. We conclude that a tetramer of integrase is associated with the viral DNA ends in the SSC.

Discussion

Retroviral DNA integration is mediated through a series of stable nucleoprotein complexes

The association of integrase with PICs isolated from virus-infected cells is highly stable. It remains tightly associated with the viral DNA even after treatment with greater than 0.5 M KCl (Lee and Craigie, 1994; Farnet and Bushman, 1997; Wei *et al*, 1997). However, it has previously not been possible to assemble and directly study such stable complexes *in vitro*. This has impeded progress in understanding many aspects

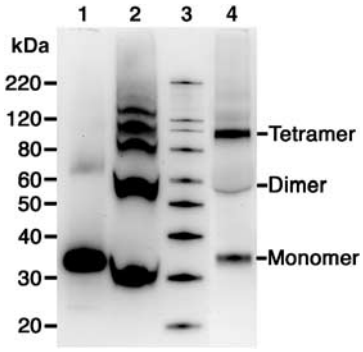


Figure 7 The SSC contains a tetramer of integrase. Complexes were assembled in the presence of inhibitor and crosslinked with DSS before separation in an agarose gel. Bands were excised and extracted protein was subjected to electrophoresis in a denaturing polyacrylamide gel. The gel was transferred to a membrane and probed for integrase by Western blotting. Lane 1, integrase without crosslinking; lane 2, integrase after crosslinking in solution; lane 3, size marker; and lane 4, integrase from crosslinked SSCs. We note that the electrophoretic mobility of the crosslinked integrase subunits is slightly faster than expected based on their molecular weight relative to the non-crosslinked size markers. This is owing to conformational constraints resulting from crosslinking. However, comparison of the ladder of integrase subunits crosslinked in solution with the crosslinked species present in the SSC unambiguously establishes that the major form is tetrameric.

of the integration mechanism because the low abundance of PICs in extracts made from infected cells makes them refractory to study by most available techniques. We show that integrase alone is able to synapse a pair of viral DNA ends and association of integrase with viral DNA in the resulting SSC is apparently as stable as the association of integrase with viral DNA in the PIC. The population of SSCs formed in the presence of inhibitor contains both processed and unprocessed viral DNA ends, but the inhibitor used to accumulate SSCs may contribute to this slow processing. However, processing does not need to be rapid in the cell because there is typically a long delay between reverse transcription and nuclear entry. Indeed, in the case of MLV, the viral DNA is initially blunt at early times after reverse transcription and the processed form accumulates over a period of several hours (Roth *et al*, 1989), although asynchrony in steps before processing may contribute to this spread. The product of concerted integration remains stably associated with integrase in our *in vitro* experiments. It seems unlikely that repair of the single strand gaps between viral and host DNA could occur without prior disassembly of the STC. Yet, for MLV, the delay between 3′ end joining and completion of 5′ end joining is estimated to be less than 1 h (Roe *et al*, 1997). Cellular proteins may be required to actively disassemble the STC as is the case for the prokaryotic transposon phage Mu where ClpX disassembles the complex (Levchenko *et al*, 1995). It is noteworthy that only integrase associated with pairs of viral DNA ends forms stable complexes and such complexes are not formed with single DNA ends or half-site reaction products. Stable complexes are formed only under reaction conditions that promote concerted DNA integration. The previously reported stabilization of the association of integrase with viral DNA in the presence of divalent metal ion (Ellison and Brown, 1994) may result from aggregation because neither SSCs or STCs are formed under these reaction conditions (data not shown).

Stable higher-order nucleoprotein complexes are involved in many transpositional DNA recombination reactions (Mizuuchi, 1992; Chaconas, 1999). In the case of bacteriophage Mu, these complexes have been called transpososomes (Surette *et al*, 1987). An early step in the Mu transposition reaction is assembly of an SSC in which the Mu ends are engaged by transposase, but not cleaved (Mizuuchi *et al*, 1992). This SSC is structurally related to the cleaved donor complex and STC that are generated by the subsequent reaction steps of donor DNA cleavage and DNA stand transfer. Thus, each of the transposition reaction steps is accompanied by changes in the organization of the transpososomes. Although retroviruses and DNA transposons differ in their lifestyles, they share the same requirement for tight regulation and control to channel the reaction along the correct pathway. The nucleoprotein complexes we identify here as intermediates in retroviral DNA integration demonstrate that this reaction parallels transpositional recombination not only in the chemical steps but also in the regulation of the reaction pathway through a series of stable nucleoprotein intermediates.

Coupling integration of pairs of viral DNA ends

Integration of viral DNA into the host genome requires that both ends are inserted at the site of integration and the half-site reaction be avoided. An early clue as to the possible regulatory mechanisms to ensure this outcome came from the finding that a mutation at one end of MLV DNA that abolishes 3′ processing also prevented processing of the 'good' end (Murphy and Goff, 1992). This suggested that there is a requirement for the pairing of ends before the reaction can be initiated and this could serve as a regulatory gate. However, different mutations that block 3′ processing at one end of HIV-1 DNA were later shown to allow processing at the 'good' end (Chen and Engelman, 2001). How can these findings be reconciled? Our results have established a series of assembly steps that provide the scaffold for the chemical steps of integration. We envision that the MLV mutation blocked the assembly of a stable complex of a pair of viral DNA ends with integrase and therefore no reaction could be initiated, whereas the different mutations in the HIV-1 experiment allowed complex assembly and therefore the 'good' end could be correctly processed. In the cell, the delay between completion of reverse transcription and transport of the PIC to the nucleus ensures that 3′ processing is usually completed before the SSC sees the target DNA. In our reactions, covalent joining of each viral DNA end to target DNA occurs sequentially with a long-lived intermediate in which only one end is joined. We suspect that this results from incomplete 3′ end processing even after one viral DNA end has been joined to target DNA; the unprocessed end must await processing before it can inserted. If this interpretation is correct, the lag between the joining of the two viral DNA ends may not reflect the kinetics *in vivo* where the PIC must await transport to the nucleus before the target DNA is encountered; both ends are usually processed by the time of entry into the nucleus. However, the slow kinetics highlight that the two joining steps can be temporally uncoupled. It is striking that essentially all the STC with only one end joined is chased to concerted integration product. Thus, once the proper complex has been correctly assembled, concerted integration is highly efficient. This is in contrast to the dead-end half-site

product, which neither forms a stable complex with integrase nor continues on the reaction pathway leading to concerted integration.

DNA end requirements for SSC assembly

The terminal DNA sequence requirements for retroviral DNA integration *in vivo* and *in vitro* are quite flexible. The CA dinucleotide that marks the boundary of integrated proviral DNA is universally important; other nucleotides that influence the efficiency of integration are few in number and generally limited to those that lie within about 10 bp of the end of the viral DNA. In *in vitro* assays with short oligonucleotide substrates that mimic the U3 or U5 ends of HIV-1 DNA, U5 ends have been reported to be significantly more active than U3 ends (Bushman and Craigie, 1991; Vink *et al*, 1991). Under our assay conditions for concerted DNA integration with long DNA substrates this preference for U5 over U3 ends is greatly diminished. In reactions containing both U5 and U3 end substrates close to 1 kb in length, the ratio of U3:U3, U3:U5 and U5:U5 products is approximately 1:2:1, as expected for random assortment and the three possible combination of pairs of viral DNA ends exhibit very similar kinetics (Supplementary Figure 2). Therefore, it is unlikely that our results are biased by forming complexes with only U5 viral DNA ends.

Although specificity is limited to the ends of the viral DNA, footprinting studies reveal that several hundred base pairs of terminal viral sequence are protected in the PIC and this protection is dependent on the presence of integrase (Wei *et al*, 1997; Chen and Engelman, 1998; Wei *et al*, 1998). Efficient concerted integration *in vitro* also requires several hundred base pairs of non-specific DNA flanking the viral terminal sequence. This flanking DNA is required for efficient assembly of the SSC (data not shown). The role of this DNA is unclear because in the SSC only about 16 bp of terminal sequence is protected in footprinting experiments. The additional DNA may facilitate assembly of the SSC through weaker non-specific interactions with additional protomers of integrase. Although the extensive protection seen in PIC footprints is dependent on the presence of integrase, it is not known if this reflects protection by integrase itself, or recruitment of other proteins by integrase.

A tetramer of integrase stably locks onto the viral DNA ends and catalyzes concerted integration

A tetramer of integrase is stably bound to the ends of the viral DNA in the SSC. The absence of crosslinked trimer in the SSC, in contrast to the products of crosslinking integrase in solution, indicates that intra-dimer and inter-dimer contacts in the tetramer, are not equivalent. The tetramer, therefore, consists of a dimer of dimers and lacks four-fold symmetry that would predict a crosslinked trimer as an intermediate in the formation of crosslinked tetramers. Although we cannot eliminate the possibility that other more loosely bound protomers participate in integration, the active sites that carry out 3′ end processing and DNA strand transfer must reside within this tightly bound tetramer that protects the viral DNA termini as demonstrated by footprinting. The extensively studied Mu transposase, which has a catalytic domain that is structurally similar to that of HIV-1 integrase (Dyda *et al*, 1994), also functions as a tetramer (Lavoie *et al*, 1991; Baker and Mizuuchi, 1992; Mizuuchi *et al*, 1992). However, a tetrameric functional complex is not a universal feature of this class of proteins because Tn5 transposase, which also has a structurally similar catalytic domain, functions as a dimer (Davies *et al*, 2000). Mixing experiments with wild-type and catalysis-defective IS10 transposase suggest that this protein also functions as a dimer (Bolland and Kleckner, 1996).

A major obstacle to obtaining detailed structural information on the active complex of HIV-1 integrase with viral DNA substrate has been the non-specific binding of integrase to DNA. The ability to form active SSCs with properties that mimic the association of integrase with viral DNA in the PIC should pave the way forward for such studies. The most promising class of integrase inhibitors for development as therapeutic agents recognize the active site in the DNA-bound configuration (Espeseth *et al*, 2000). High-resolution structural information on the HIV-1 integrase-active site engaged with viral DNA and such inhibitors would be invaluable for the development of inhibitors that target HIV-1 integrase.

Materials and methods

DNA substrates

The 1513 bp linear mini-viral DNA was excised from pSca355 plasmid (Li and Craigie, 2005), which is based on pCR2.1 (Invitrogen), by cutting with *Sca*I and *Hinc*II and purified by agarose gel electrophoresis. Digestion of the mini-viral DNA by *Sca*I results in a linear fragment of plasmid sequence flanked by 31 bp of HIV-1 NL4-3 blunt-end U5 LTR sequence. Shorter substrate DNAs were prepared by further digestion with different restriction enzymes. The DNA fragments were treated with alkaline phosphatase and then 5′ end labeled with [γ-^{32}P]ATP by T4 polynucleotide kinase. For DNase I footprint studies, the 5′-end-labeled 1513 bp viral DNA was digested with *Ban*I to produce a 984 bp single-end-labeled DNA. To make internally labeled substrate DNA, a 950 bp DNA was excised from pSca355 by cleavage with *Eco*RI and *Ban*I, purified by agarose gel electrophoresis and then ligated at the *Eco*RI site to a duplex 35 bp ^{32}P-labeled DNA corresponding to 31 bp of U5 LTR terminal sequence. The ligated 985 bp donor DNA was finally purified by agarose gel electrophoresis. The processed viral DNA end with a 3′ ddA was made by transferring ddATP to the oligo 5′ pAATTCTTTTAGTCAGTGTGGAAAATCTCTAGC using TdT and annealing with the 5′-^{32}P-labeled oligo 5′ pACTGCTAGAGATTTTCC ACACTGACTAAAAG. The annealed duplex was ligated with with a 970 bp *Eco*R1/*Pst*1 fragment of pCR2.1 and purified by agarose gel electrophoresis.

Protein expression and purification

His-tagged HIV-1 integrase W235F was purified as described (Li and Craigie, 2005). The W235F mutation improves the efficiency of the concerted integration reaction *in vitro* (Li and Craigie, 2005) and does not compromise viral replication *in vivo* (Leavitt *et al*, 1996). The W235F mutation improves both the efficiency of concerted integration and the efficiency of formation of the complexes reported in this paper by a factor of approximately two-fold over wild-type HIV-1 integrase (Supplementary Figure 1). Briefly, the integrase was expressed in *Escherichia coli* BL21(DE3) and the cells were lysed in buffer containing 0.1 M NaCl. The lysate was centrifuged and integrase was extracted from the pellet in buffer containing 2 M NaCl. The protein was then purified by nickel-affinity chromatography and the His tag was removed with thrombin.

Integration assay and complex assembly

Typical reaction mixtures (25 μl final volume) were assembled by incubating 80 nM integrase on ice in 20 mM HEPES, pH 7.5, 12% DMSO, 5 mM DTT, 10% PEG-6000, 10 mM $MgCl_2$, 20 μM $ZnCl_2$ and 100 mM NaCl (final), followed by addition of 10 nM viral DNA substrate. These components were preincubated on ice for 0.5 h and 250 ng target plasmid DNA pBR322 was then added. After a further 0.5 h preincubation on ice, the reaction was initiated by transfer to

37°C and incubation was continued for 1 h, or as noted otherwise. The reactions were stopped by addition of SDS and EDTA to 0.1% and 10 mM, respectively, together with 5 μg of proteinase K. Incubation was continued at 37°C for a further 1 h. A 2.5 μl aliquot of the reaction mixture was then subjected to electrophoresis in a 0.8% agarose gel (12 × 24 cm) in 1 × TBE buffer. Gels were dried, exposed to imaging plates and visualized and quantified with a Fuji BAS-2500 bio-imaging analyzer. Complex assembly was carried out in the same way, except that where stated the integrase was preincubated with 500 μM inhibitor 11 (Zhang *et al*, 2003). EDTA was added to 10 mM and BSA was added to 1 mg/ml before electrophoresis of complexes.

Electrophoresis of complexes and gel isolation of products

A 2.5 μl aliquot of sample was subjected to electrophoresis in an 0.8% agarose gel in TBE buffer with 1 M urea at 7.7 V/cm for 3.5 h at 4°C. For isolation of complexes, markers were run side by side with the radiolabeled complexes and bands were cut out after staining with ethidium bromide. For two-dimensional gel electrophoresis, 2.5 μl of reaction mixture was first subjected to electrophoresis in a 0.8% agarose gel containing 1 M urea (10 × 15 cm) in TBE buffer with 1 M urea at 7.7 V/cm at 4°C. After the first dimension of native electrophoresis, the gel was soaked in 200 ml 1 × TBE with 0.5% SDS for 30 min, followed by 200 ml 1 × TBE with 0.1% SDS for a further 30 min. The gel was then rotated 90° and subjected to electrophoresis in 1 × TBE buffer with 0.1% SDS at 7.7 V/cm at room temperature for about 1 h. The gel was then dried and exposed.

DNase I footprinting

A one-end labeled 985 bp donor DNA (blunt end) or a 1005 bp processed donor DNA with a 3′ ddA was used to assemble SSCs for DNase I protection. After assembly of synaptic complexes at 37°C for 2 h in the presence of 500 μM inhibitor 11 (blunt-end substrate) or absence of inhibitor (processed substrate), the mixture was cooled to room temperature and 0.01 U DNase I (New England Biolabs) was added. After incubation at room temperature for 2 min, digestion was stopped by addition of EDTA to 10 mM. The complexes were then separated on a urea–TBE gel as described above and DNA from the excised complexes was recovered by the 'crush and soak' method (Sambrook *et al*, 1989). Equivalent amounts of recovered DNA from each band were solubilized in formamide loading buffer and subjected to denaturing 15% PAGE. The dried gels were exposed for 1 week and analyzed by a Fuji BAS2500 bio-imaging analyzer. DNA markers for footprints were prepared by the Maxam–Gilbert method (Maxam and Gilbert, 1977).

Chemical crosslinking and Western blotting

Chemical crosslinking of IN or complexes was performed using DSS, a homobifunctional chemical crosslinker with an 11.4-Å spacer arm. DSS was freshly dissolved at 25 mM in dimethyl sulfoxide. The reaction was initiated by the addition of 0.5 mM DSS to IN alone or 0.8 mM DSS to complex assembly reactions. Incubation was for 30 min at room temperature. The reaction was stopped by the addition of 1 M Tris–HCl (pH 8.0) to 50 mM, and kept on ice for a further 15 min. Complexes were subjected to electrophoresis in urea–TBE agarose gels, excised and eluted in 1 × SDS–PAGE running buffer with 0.1 mg/ml insulin as carrier. The eluted protein was precipitated by TCA-deoxycholate (DOC)/Acetone. Briefly, 1/100 volume of 2% Na DOC was added and kept at room temperature for 15 min, followed by the addition of 1/4 volume of 100% (w/v) cold TCA. After incubation for 30 min on ice, the precipitate was centrifuged at 16 000 *g* for 20 min and washed twice with ice-cold acetone. The dried pellets were resuspended and subjected to electrophoresis in a 4–12% Bis–Tis gel (Invitrogen). Western blotting was carried out using an Amplified Alkaline Phosphatase Goat Anti-Rabbit Immuno-Blot Assay Kit (Bio-Rad). Briefly, the gel was blotted to PVDF membrane (Invitrogen) and the membrane was blocked overnight at 4°C. It was then incubated with 5 ml of 1:100 diluted integrase antiserum overnight at 4°C, followed by addition of biotinylated goat anti-rabbit antibody and streptavidin-biotinylated alkaline phosphatase complex. Detection was carried out with ECL+ using an LAS-1000 (Fuji). The rabbit polyclonal anti-HIV-1 integrase antiserum no. 756 (Grandgenett and Goodarzi, 1994) was from the NIH AIDS Research and Reference Reagent Program and a gift from Dr Duane P Grandgenett.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online.

Acknowledgements

This research was supported in part by the Intramural Research Program of the NIH, National Institute of Diabetes and Digestive and Kidney Diseases, National Cancer Institute and by the NIH AIDS Targeted Antiviral Program. We thank Kiyoshi Mizuuchi for valuable discussions during the course of this work and Kiyoshi Mizuuchi, Wei Yang and members of the Craigie laboratory for their comments on the manuscript. The following reagent was obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: Antiserum to HIV-1 Integrase (no. 756) from Dr Duane P Grandgenett.

References

Baker TA, Mizuuchi K (1992) DNA-promoted assembly of the active tetramer of the Mu transposase. *Genes Dev* **6:** 2221–2232

Bao KK, Wang H, Miller JK, Erie DA, Skalka AM, Wong I (2003) Functional oligomeric state of avian sarcoma virus integrase. *J Biol Chem* **278:** 1323–1327

Bolland S, Kleckner N (1996) The three chemical steps of Tn10/IS10 transposition involve repeated utilization of a single active site. *Cell* **84:** 223–233

Bowerman B, Brown PO, Bishop JM, Varmus HE (1989) A nucleoprotein complex mediates the integration of retroviral DNA. *Genes Dev* **3:** 469–478

Brown PO, Bowerman B, Varmus HE, Bishop JM (1987) Correct integration of retroviral DNA *in vitro*. *Cell* **49:** 347–356

Brown PO (1997) Integration. In *Retroviruses*, Coffin, JM, Hughes SH, Varmus HE (eds) pp 161–203. New York: Cold Spring Harbor Laboratory Press

Bujacz G, Jaskolski M, Alexandratos J, Wlodawer A, Merkel G, Katz RA, Skalka AM (1995) High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J Mol Biol* **253:** 333–346

Bushman FD, Craigie R (1991) Activities of human immunodeficiency virus (HIV) integration protein *in vitro*: Specific cleavage and integration of HIV DNA. *Proc Natl Acad Sci USA* **88:** 1339–1343

Bushman FD, Fujiwara T, Craigie R (1990) Retroviral DNA integration directed by HIV integration protein *in vitro*. *Science* **249:** 1555–1558

Cai M, Zheng R, Caffrey M, Craigie R, Clore GM, Gronenborn AM (1997) Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat Struct Biol* **4:** 567–577

Chaconas G (1999) Studies on a 'jumping gene machine': higher-order nucleoprotein complexes in Mu DNA transposition. *Biochem Cell Biol* **77:** 487–492

Chaconas G, Harshey RM (2002) Transposition of phage Mu DNA. In *Mobile DNA II*, Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) pp 384–402. Washington, DC: ASM Press

Chen HM, Engelman A (1998) The barrier-to-autointegration protein is a host factor for HIV type 1 integration. *Proc Natl Acad Sci USA* **95:** 15270–15274

Chen HM, Engelman A (2001) Asymmetric processing of human immunodeficiency virus type 1 cDNA *in vivo*: implications for functional end coupling during the chemical steps of DNA transposition. *Mol Cell Biol* **21:** 6758–6767

Chen JCH, Krucinski J, Miercke LJW, Finer-Moore JS, Tang AH, Leavitt AD, Stroud RM (2000a) Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc Natl Acad Sci USA* **97:** 8233–8238

Chen ZG, Yan YW, Munshi S, Li Y, Zugay-Murphy J, Xu B, Witmer M, Felock P, Wolfe A, Sardana V, Emini EA, Hazuda D, Kuo LC (2000b) X-ray structure of simian immunodeficiency virus integrase containing the core and C-terminal domain (residues 50–293)—an initial glance of the viral DNA binding platform. *J Mol Biol* **296:** 521–533

Craigie R (2002) Retroviral DNA integration. In *Mobile DNA II*, Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) pp 613–630. Washington, DC: ASM Press

Davies DR, Goryshin IY, Reznikoff WS, Rayment I (2000) Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* **289:** 77–85

Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* **266:** 1981–1986

Ellison V, Abrams H, Roe T, Lifson J, Brown PO (1990) Human immunodeficiency virus integration in a cell-free system. *J Virol* **64:** 2711–2715

Ellison V., Brown PO (1994) A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration *in vitro*. *Proc Natl Acad Sci USA* **91:** 7316–7320

Engelman A, Mizuuchi K, Craigie R (1991) HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* **67:** 1211–1221

Espeseth AS, Felock P, Wolfe A, Witmer M, Grobler J, Anthony N, Egbertson M, Melamed JY, Young S, Hamill T, Cole JL, Hazuda D (2000) HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase. *Proc Natl Acad Sci USA* **97:** 11244–11249

Farnet CM, Bushman FD (1997) HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes *in vitro*. *Cell* **88:** 483–492

Farnet CM, Haseltine WA (1990) Integration of human immunodeficiency virus type 1 DNA *in vitro*. *Proc Natl Acad Sci USA* **87:** 4164–4168

Faure A, Calmels C, Desjobert C, Castroviejo M, Caumont-Sarcos A, Tarrago-Litvak L, Litvak S, Parissi V (2005) HIV-1 integrase crosslinked oligomers are active *in vitro*. *Nucleic Acids Res* **33:** 977–986

Grandgenett DP, Goodarzi G (1994) Folding of the multidomain human immunodeficiency virus type-I integrase. *Protein Sci* **3:** 888–897

Hindmarsh P, Ridky T, Reeves R, Andrake M, Skalka AM, Leis J (1999) HMG protein family members stimulate human immunodeficiency virus type 1 and avian sarcoma virus concerted DNA integration *in vitro*. *J Virol* **73:** 2994–3003

Lavoie BD, Chan BS, Allison RG, Chaconas G (1991) Structural aspects of a higher-order nucleoprotein complex—induction of an altered DNA–structure at the Mu–host junction of the Mu-Type-1 transpososome. *EMBO J* **10:** 3051–3059

Leavitt AD, Robles G, Alesandro N, Varmus HE (1996) Human immunodeficiency virus type 1 integrase mutants retain *in vitro* integrase activity yet fail to integrate viral DNA efficiently during infection. *J Virol* **70:** 721–728

Lee MS, Craigie R (1994) Protection of retroviral DNA from autointegration: involvement of a cellular factor. *Proc Natl Acad Sci USA* **91:** 9823–9827

Levchenko I, Luo L, Baker TA (1995) Disassembly of the Mu transposase tetramer by the ClpX *chaperone*. *Genes Dev* **9:** 2399–2408

Li M, Craigie R (2005) Processing the viral DNA ends channels the HIV-1 integration reaction to concerted integration. *J Biol Chem* **280:** 29334–29339

Lodi PJ, Ernst JA, Kuszewski J, Hickman AB, Engelman A, Craigie R, Clore GM, Gronenborn AM (1995) Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry* **34:** 9826–9833

Maxam AM, Gilbert W (1977) New method for sequencing DNA. *Proc Natl Acad Sci USA* **74:** 560–564

Mizuuchi K (1992) Transpositional recombination—mechanistic insights from studies of Mu and other elements. *Annu Rev Biochem* **61:** 1011–1051

Mizuuchi K (1997) Polynucleotidyl transfer reactions in site-specific DNA recombination. *Genes Cells* **2:** 1–12

Mizuuchi M, Baker TA, Mizuuchi K (1992) Assembly of the active form of the transposase–Mu DNA complex: a critical control point in Mu transposition. *Cell* **70:** 303–311

Murphy JE, Goff SP (1992) A mutation at one end of Moloney murine leukemia virus DNA blocks cleavage of both ends by the viral integrase *in vivo*. *J Virol* **66:** 5092–5095

Pommier Y, Johnson A, Marchand C (2005) Integrase inhibitors to treat HIV/AID. *Nat Rev Drug Disc* **4:** 236–248

Roe T, Chow SA, Brown PO (1997) 3′-End processing and kinetics of 5′-end joining during retroviral integration *in vivo*. *J Virol* **71:** 1334–1340

Roth MJ, Schwartzberg PL, Goff SP (1989) Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence on IN function and terminal DNA sequence. *Cell* **58:** 47–54

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Press

Sherman PA, Fyfe JA (1990) Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selective DNA cleaving activity. *Proc Natl Acad Sci USA* **87:** 5119–5123

Sinha S, Grandgenett DP (2005) Recombinant human immunodeficiency virus type 1 integrase exhibits a capacity for full-site integration *in vitro* that is comparable to that of purified preintegration complexes from virus-infected cells. *J Virol* **79:** 8208–8216

Sinha S, Pursley MH, Grandgenett DP (2002) Efficient concerted integration by recombinant human immunodeficiency virus type 1 integrase without cellular or viral cofactors. *J Virol* **76:** 3105–3113

Surette MG, Buch SJ, Chaconas G (1987) Transpososomes—stable protein–DNA complexes involved in the *in vitro* transposition of bacteriophage Mu DNA. *Cell* **49:** 253–262

Vink C, van Gent DC, Elsgerma Y, Plasterk RHA (1991) Human immunodeficiency virus integrase protein requires a subterminal position of its viral DNA. *J. Virol*. recognition sequence for efficient cleavage. *J Virol* **65:** 4636–4644

Vora A, Grandgenett DP (2001) DNase protection analysis of retrovirus integrase at the viral DNA ends for full-site integration *in vitro*. *J Virol* **75:** 3556–3567

Wang JY, Ling H, Yang W, Craigie R (2001) Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J* **20:** 7333–7343

Wei SQ, Mizuuchi K, Craigie R (1997) A large nucleoprotein assembly at the ends of the viral DNA mediates retroviral DNA integration. *EMBO J* **16:** 7511–7520

Wei SQ, Mizuuchi K, Craigie R (1998) Footprints on the viral DNA ends in Moloney murine leukemia virus preintegration complexes reflect a specific association with integrase. *Proc Natl Acad Sci USA* **95:** 10535–10540

Yang ZN, Mueser TC, Bushman FD, Hyde CC (2000) Crystal structure of an active two-domain derivative of rous sarcoma virus integrase. *J Mol Biol* **296:** 535–548

Zhang XC, Pais GCG, Svarovskaia ES, Marchand C, Johnson AA, Karki RG, Nicklaus MC, Pathak VK, Pommier Y, Burke Jr TR (2003) Azido-containing aryl beta-diketo acid HIV-1 integrase inhibitors. *Bioorg Med Chem Lett* **13:** 1215–1219

